

PATENT APPLICATION
Docket No. 2852.2.1

UNITED STATES PATENT APPLICATION

of

David C. Taylor

and

Ronald L. Hall, Jr.

for

USER-CONTEXT-BASED SEARCH ENGINE

MADSON & METCALF, P.C.
ATTORNEYS AT LAW
900 GATEWAY TOWER WEST
15 WEST SOUTH TEMPLE
SALT LAKE CITY, UTAH 84101

Year	1900	1901	1902	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
Population	1,000,000	1,050,000	1,100,000	1,150,000	1,200,000	1,250,000	1,300,000	1,350,000	1,400,000	1,450,000	1,500,000	1,550,000	1,600,000	1,650,000	1,700,000	1,750,000	1,800,000	1,850,000	1,900,000	1,950,000	2,000,000	2,050,000	2,100,000	2,150,000	2,200,000	2,250,000	2,300,000	2,350,000	2,400,000	2,450,000	2,500,000	2,550,000	2,600,000	2,650,000	2,700,000	2,750,000	2,800,000	2,850,000	2,900,000	2,950,000	3,000,000	3,050,000	3,100,000	3,150,000	3,200,000	3,250,000	3,300,000	3,350,000	3,400,000	3,450,000	3,500,000	3,550,000	3,600,000	3,650,000	3,700,000	3,750,000	3,800,000	3,850,000	3,900,000	3,950,000	4,000,000	4,050,000	4,100,000	4,150,000	4,200,000	4,250,000	4,300,000	4,350,000	4,400,000	4,450,000	4,500,000	4,550,000	4,600,000	4,650,000	4,700,000	4,750,000	4,800,000	4,850,000	4,900,000	4,950,000	5,000,000	5,050,000	5,100,000	5,150,000	5,200,000	5,250,000	5,300,000	5,350,000	5,400,000	5,450,000	5,500,000	5,550,000	5,600,000	5,650,000	5,700,000	5,750,000	5,800,000	5,850,000	5,900,000	5,950,000	6,000,000	6,050,000	6,100,000	6,150,000	6,200,000	6,250,000	6,300,000	6,350,000	6,400,000	6,450,000	6,500,000	6,550,000	6,600,000	6,650,000	6,700,000	6,750,000	6,800,000	6,850,000	6,900,000	6,950,000	7,000,000	7,050,000	7,100,000	7,150,000	7,200,000	7,250,000	7,300,000	7,350,000	7,400,000	7,450,000	7,500,000	7,550,000	7,600,000	7,650,000	7,700,000	7,750,000	7,80																																																																

USER-CONTEXT-BASED SEARCH ENGINE

BACKGROUND OF THE INVENTION

1. Related U.S. Applications

This application claims priority to U.S. Provisional Patent Application Serial No. 60/146,878, filed August 3, 1999, entitled "Data Extraction Tool," which application is incorporated herein by reference.

2. The Field of the Invention

This invention relates to a data extraction tool and, more particularly, to novel systems and methods for organizing and presenting information about products for sale over the Internet for a user.

3. The Relevant Technology

In what is known as the information age, information is readily available electronically, through information repositories known as datastores and databases. Datastores are substantially unorganized collections of data, while databases are indexed in some fashion. The Internet, the world's largest database, has made available enormous quantities of information to anyone with a personal computer and Internet access. This can be very helpful for people who wish to learn about something or conduct business in the convenience of their own homes. However, it can also be tremendously time-consuming to locate a desired bundle of information among the millions available.

The Internet is organized only by the name of each web site. Each individual or group maintaining a web site decides how that web site will be organized. Thus, there is no official catalog of information available on the Internet. Anyone desiring information must hypothesize which web sites would be likely to have the desired data and navigate through

Moreover, information obtained through a search often becomes outdated. Currently, a user must revisit previously found sites to determine whether the old information

1 is still valid. Additionally, a user must perform a new search to locate any newly relevant sites
2 and search through those sites for relevant information.

3 Directories function differently than search engines. Rather than search based on
4 keywords provided by a user, most directories provide a user with an information scheme,
5 often hierarchically organized. The user then chooses what type of information to search for,
6 designating narrower groups of information with each choice. Ultimately, the user reaches
7 the bottom level of the hierarchy and receives a list of links to information within that level.

8 Directories are advantageous in that information concerning a certain topic is
9 typically grouped together. A directory probably will not inundate a user with information,
10 but rather provide a few links believed to be important by the creators of the directory.
11 Nevertheless, directories have drawbacks of their own.

12 For example, traditional directories contain information deemed of value by those
13 who compile them. A user may have an entirely different view of what is important and what
14 is irrelevant. A user may thus find that information he or she needs simply is not available
15 on the directory.

16 Also, directories take time to navigate. A user must make a series of decisions to
17 reach any useful information at all. Even then, a user may find it necessary to backtrack and
18 choose a different route through the hierarchy. Since a user cannot fashion groupings of
19 information, he or she may be required to view several branches of the hierarchy to obtain the
20 full range of information he or she desires.

21 Moreover, if a user does not know how to classify the bit of information sought, he
22 or she may not even be able to find it in the directory. For example, a user desiring to find
23 the meaning of "salmonella" in a biological directory may spend great amounts of time
24 looking through the "aquatic life" branch of the directory, without ever realizing that
25 "salmonella" is more properly classified as "microscopic life." The more a user's view of how
26

1 information should be organized differs form that of the directory's creators, the more
2 difficult it will be for the user to find information in the directory.

3 Consequently, there is a need for a data extraction tool capable of providing many
4 of the benefits of both search engines and directories, without drawbacks listed above. For
5 example, there is a need for a tool that could reliably provide a list of highly relevant
6 information locations based on a simple text query. Furthermore, such a tool should provide
7 ready access to the exact location of the information. Preferably, the tool would supply the
8 user with a list of locations or links that can be easily sorted and updated for the convenience
9 of the user. Furthermore, the tool should not require that the user understand the
10 configuration of the tool's internal databases.

11 In addition to the problems mentioned above, current searching methods are deficient
12 in a number of other ways. Consequently, a more advanced data extraction tool may provide
13 numerous benefits to those desiring to obtain information from a large datastore or database,
14 such as the Internet.

15 16 OBJECTS AND BRIEF SUMMARY OF THE INVENTION

17 In view of the foregoing, it is a primary object of the present invention to provide a
18 data extraction tool capable of context-sensitive searching, pinpointing, databasing,
19 automatically updating information, or any combination thereof, from a datastore or database.
20 Consistent with the foregoing object, and in accordance with the invention as embodied and
21 broadly described herein, an apparatus and method are disclosed, in suitable detail to enable
22 one of ordinary skill in the art to make and use the invention. In certain embodiments, an
23 apparatus and method in accordance with the present invention may be directed to locating
24 prices for products that can be purchased over the Internet.

25 Selected embodiments of the invention have modules for mining, databasing,
26 searching, filtering, pinpointing, presenting, indexing, or updating information for a user, or

1 some combination thereof. According to some embodiments, the data extraction tool mines
2 information from the information source and organizes the locations of that information
3 within a local database. Then, a user may query the tool for a desired type of information.
4 The tool filters the local database to provide a set of pinpoint site locations with information
5 of the type requested in the query. These pinpoint site locations are presented to a user and
6 indexed for future reference. The index of site locations may be updated automatically by the
7 tool.

8 A context system is provided for manually or automatically determining the proper
9 context for a user's query. Thus, the data extraction tool provides information with a high
10 probability of relevance to the user. The user obtains the information without expending
11 much effort to refine the search.

12 These and other objects, features, and advantages of the present invention will
13 become more fully apparent from the following description and appended claims, or may be
14 learned by the practice of the invention as set forth hereinafter.

15 BRIEF DESCRIPTION OF THE DRAWINGS

16 In order that the manner in which the above-recited and other advantages and objects
17 of the invention are obtained will be readily understood, a more particular description of the
18 invention briefly described above will be rendered by reference to specific embodiments
19 thereof which are illustrated in the appended drawings. Understanding that these drawings
20 depict only typical embodiments of methods and apparatus in accordance with the invention
21 and are not therefore to be considered to be limiting of its scope, the invention will be
22 described and explained with additional specificity and detail through the use of the
23 accompanying drawings in which:
24

25 Figure 1 is a schematic block diagram of an individual node in a network system;
26

1 Figure 2 is a schematic block diagram of one embodiment of a data extraction tool
2 on a server, such as the server of Figure 1;

3 Figure 3 is a schematic block diagram of a data extraction tool, as shown in Figure
4 2, configured for use with a node in a network system, as shown in Figure 1;

5 Figure 4 is a schematic block diagram of data structures for administering and
6 executing a user interface in accordance with the invention;

7 Figure 5 is a schematic block diagram of data structures for administering and
8 executing a filtering module in accordance with the invention;

9 Figure 6 is a schematic block diagram of data structures for administering and
10 executing an attributes index in accordance with the invention;

11 Figure 7 is a schematic block diagram of methods for implementing one embodiment
12 of the data structures and functions of Figure 2 in accordance with the invention;

13 Figure 8 is a schematic block diagram of methods for implementing one embodiment
14 of the mining step of Figure 7 in accordance with the invention;

15 Figure 9 is a schematic block diagram of methods for implementing one embodiment
16 of the database construction step of Figure 7 in accordance with the invention;

17 Figure 10 is a schematic block diagram of methods for implementing one embodiment
18 of the searching step of Figure 2 in accordance with the invention;

19 Figure 11 is a schematic block diagram of an alternative method for implementing
20 the data structures and functions of Figure 2;

21 Figure 12 is a schematic block diagram of methods for implementing one embodiment
22 of the context construction module of Figure 11 in accordance with the invention;

23 Figure 13 is a schematic block diagram of methods for implementing one embodiment
24 of the context comparison module of Figure 11 in accordance with the invention;

25 Figure 14 is a schematic block diagram of methods for implementing one embodiment
26 of the information matching module of Figure 11 in accordance with the invention;

1 Figure 15 is a schematic diagram of a hierarchical database usable in conjunction with
2 the present invention;

3 Figure 16 is a schematic block diagram of a possible embodiment of a user interface
4 according to the present invention; and

5 Figure 17 is an exemplary information chart depicting how information may be
6 presented to a user for modification and rearrangement, in accordance with the invention.

7
8 **DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS**

9 The presently preferred embodiments of the present invention will be best understood
10 by reference to the drawings, wherein like parts are designated by like numerals throughout.
11 It will be readily understood that the components of the present invention, as generally
12 described and illustrated in the figures herein, could be arranged and designed in a wide
13 variety of different configurations. Thus, the following more detailed description of the
14 embodiments of the apparatus, system, and method of the present invention, as represented
15 in figures 1 through 17, is not intended to limit the scope of the invention, as claimed, but is
16 merely representative of presently preferred embodiments of the invention.

17 Those of ordinary skill in the art will, of course, appreciate that various modifications
18 to the details of the figures may easily be made without departing from the essential
19 characteristics of the invention. Thus, the following description of the figures is intended only
20 by way of example, and simply illustrates certain presently preferred embodiment consistent
21 with the invention as claimed.

22 Referring now to Figure 1, a system 10 or network 10, such as the Internet, may
23 include nodes 11 (e.g. nodes 50, 52, 54). Each node 11 may include a processor 12 and
24 memory devices 14, such as storage devices 16, read only memory (ROM) 18, and random
25 access memory (RAM) 20, sometimes referred to as operational memory. The node 11 may
26

1 include a variety of input devices 22, and output devices 24 whether dedicated as illustrated
2 in Figure 1, or more generally available over a network.

3 Typically, a node 11 may include a network card 26 for connecting to a network 30
4 (e.g. network 10) outwardly, and a bus 32 for interconnecting elements internally.

5 Input devices 22 may include a keyboard 34, a mouse 36 or other pointing device 36
6 such as a stylus or graphics tablet, an interactive touch screen 38, a scanner 40, or even a
7 storage device 41 for providing data to the node 11. Similarly, output devices 24 may include
8 monitor 42, printer 44, storage devices 46, and the like for providing data from the node 11.

9 A router 48 may interconnect networks 30, 50 where each network 30, 50 may
10 include some simple nodes 52, such as clients 52a-52d, and servers 54. Networks 30, 50 are
11 well understood in the art. Accordingly, the hardware illustrated is by way of example, and
12 not limitation as to the hardware suite on which the invention may be implemented. More or
13 less equipment may be used in many particular embodiments.

14 The system 10 is the datastore or database from which information is to be obtained.
15 However, the system 10 need not be configured as shown in Figure 1. For example, the
16 system 10 may be a datastore or database contained on a single computer. However, many
17 of the subsequent descriptions will refer to the system 10 as a distributed network 10 of
18 computers, such as the Internet.

19 Figure 2 shows one embodiment of a data extraction tool 110, or tool 110, with its
20 associated modules. A mining module 112 gathers information from a data source, preferably
21 the Internet. A databasing module 114 categorizes and sorts information within a local
22 database. This information can be actual data directly from the data source, or it can be
23 simply pointers to locations of data within the data source.

24 An input module 116 interfaces with a user to receive a request for information. A
25 filtering module 118 filters information to isolate the data most relevant to a user's request.
26 A pinpointing module 120 locates and returns identification of the exact location of

1 information. A presentation module 122 presents information summaries and locations to a
2 user. An indexing module 124 organizes information for use and access by a user. An
3 updating module 126 automatically updates information in a local database.

4 The arrows in Figure 2 show a general chronological flow. However, the modules
5 shown do not have to be accessed in the order shown. In addition, modules can operate
6 multiple times. For example, the updating module 126 may utilize the mining module 112 to
7 obtain more information, which may then be shown to a user by the presentation module 122
8 and stored by the indexing module 124.

9 Referring to Figure 3, a node 11 may have a hard disk 128 or HD 128, an
10 input/output port 130 or I/O 130, a central processing unit 132 or CPU 132, and a memory
11 133. The modules 112, 114, 116, 118, 120, 122, 124, and 126 may be temporarily stored for
12 use in the memory 133, permanently stored in the hard disk 128, and processed through the
13 central processing unit 132. A user interface 134 and a server link 136 provide for
14 communication with a user and with the network 10 via the I/O 130. A transaction interface
15 138 may also be included to permit purchasing and selling over the network 10.

16 Figure 4 shows some data structures that may be included in the user interface 134.
17 A home page 140 provides a user with access to the node 11 with the data extraction tool
18 110. A web server 142 stores data that makes up the home page 140. CGI scripts 144 may
19 display data from the web server 142 for a user in the form of text 146 and graphics 148.

20 A free form input module 150 receives searching parameters, in the form of a query,
21 from a user. A semantic analysis module 152 parses the query and uses context templates 154
22 to develop a list of contexts that may correlate to information desired by a user. An inquiry
23 module 156 prompts a user for further guidance concerning which of the available contexts
24 are truly relevant. A query modification module 158 modifies the query to suit a user's
25 response to prompting from the inquiry module 156. A presentation module 160 displays
26

1 search results in a clear fashion for a user, and can also permit a user to interact with other
2 nodes (e.g. nodes 50, 52, 54) where further information may be stored.

3 In addition, a site interaction module 162 can permit partial processing of information
4 by the data extraction tool 110 before presentation to a user. A pinpoint selection module
5 164 chooses relevant sites for further processing. A login module 166, if needed, may permit
6 the site interaction module 162 to automatically log onto a site where relevant information
7 is stored. A link selection module 168 chooses the most relevant path within the site for
8 retrieval of the desired information. A page parsing module 170 determines whether text
9 from the site is relevant to a user's query.

10 Referring to Figure 5, some data structures that may be used in the filtering module
11 118 are shown. The semantic net 174 is a resource for matching query text from a user to
12 text from a web site. Context clues 176 provide information for contextual comparisons
13 based on classifications 178 of contexts in which a word may be found. A context selector
14 180 selects those contexts that correlate to the proper context for the query and isolates them
15 via filters 182. The filters 182 may reference the context system 184, which simply provides
16 a list of actions corresponding to each instance of a word. For example, the context system
17 may specify that a site should be retained if a keyword is found in a certain context within the
18 site, but that the site should be filtered out if the keyword is used in a different context.

19 Referring to Figure 6, the context clues 176 and the context selector 180 may
20 reference an attributes index 185. The attributes index 185 contains a list of words 186.
21 Each word has at least one meaning 187 indexed to that word, and each meaning 187 has a
22 list of relations 188, such as synonyms, antonyms, subsets, supersets, usage correlation, and
23 usage association. A second meaning 189, and however many meanings exist for the word
24 186, may also be included with an associated list of relations.

25 The relations 188 provide context clues 176 so that a given web site can be classified
26 by context. The context may be determined, for example, by the frequency and combination

1 of relations 188 that appear within the web site. Thus, the filters 182 can filter out those web
2 sites in which the proper keyword is used in an irrelevant context.

3 Figure 7 shows one method for implementation of the data structures of Figure 2.
4 In a mining step 190, a data extraction tool 110 mines information from the network 10. The
5 locations of the information, and possibly some of the information itself, may be compiled and
6 formatted in a database construction step 191. A searching step 192 permits a user to query
7 for information stored by the database construction step 191. A filtering step 193 selects the
8 information most relevant to a user's query. A pinpointing step 194 determines the exact
9 location of the relevant information on the network 10. A presenting step 196 organizes
10 relevant information and provides it to a user. An indexing step 198 links relevant
11 information to the location of that information on the network 10. An updating step 200
12 subsequently performs an automatic search of the network 10 for new information relevant
13 to the user's query.

14 Figure 8 shows possible steps that might be included within the mining step 190. In
15 a route selection step 202, the tool 110 chooses an orderly method for processing information
16 from the network 10. Preferably, the route selection step 202 involves an orderly progression
17 to ensure that each potentially relevant parcel of data is processed once and only once. In an
18 autonavigation step 204, the tool 110 receives information from the network 10 for
19 processing in a content reading step 206.

20 In an evaluation step 208, the tool 110 evaluates the potential relevance of the text
21 146 of a site to future queries of a user. The tool 110 may be directed towards acquiring a
22 certain type of information, or broadly used to obtain and categorize a wide variety of data.
23 The scope of data to be mined determines how selective the evaluation step 110 will be. In
24 a content extraction step 210, potentially relevant content is compared against a listing of
25 needed information to further filter it in a database filtration step 212. The data are indexed
26 for ready access by an addition to a master index step 214.

1 Figure 9 shows possible steps that might be included within the database construction
2 step 191. A database structuring step 216 provides the structure and organization for the
3 information. In a schema provision step 218, a relations recording step 220 and an indices
4 recording step 222 organize data into fields that are appropriately linked together and indexed
5 for rapid reference. In an input data step 224, the tool 110 receives information gathered
6 during the mining step 186.

7 A data classification step 226 uses discrimination functions 228 to categorize
8 information within the schema developed by the schema provision step 218. A schema
9 refining step 230 permits revision of the schema as needed to accommodate information that
10 otherwise cannot be appropriately categorized within the schema. In a records filling step
11 232, the tool 110 adds data to form records.

12 In an addition to database step 234, the tool 110 adds data retrieved by the mining
13 step 186 to the local database. This involves a number of steps. In a site identification step
14 236, the tool 110 identifies sites of relevant information. In a site isolation step 238, the tool
15 110 further filters sites based on criteria provided by a user or by the programming of the tool
16 110. For example, the tool 110 can be programmed to isolate sites capable of conducting
17 commerce over the Internet. In that case, the site isolation step 238 would filter out all sites
18 without a method for conducting commerce through the site. In a site contents classification
19 step 240, the tool 110 classifies data into appropriate categories, as laid out in the schema.
20 A data selection step 242 chooses classifiable data for transmission to a record preparation
21 step 244, where data is added to records in the local database.

22 Referring to Figure 10, a number of steps may be included within the searching step
23 192. A user may request information by entering free form text or other query inputs in a
24 query receiving step 246. In a query parsing step 248, the query is compared against a list
25 of possible contexts by a semantic net reference step 250. In an inquiry preparation step 252,
26 the tool 110 forms a question for a user, in a question selection step 254, to ask for

1 clarification concerning which of the potential contexts that may match the query is the most
2 relevant.

3 The inquiry computation step 256 may provide an estimate of the time required to
4 perform a search for each potential context, so that a user will know how long the tool 110
5 will take to process a search for a given context. This is especially helpful when a user has
6 provided a very broad query. In such a case, the computation time will be high, so a user will
7 know that the search will take a comparatively long time and provide a comparatively large
8 amount of information, perhaps more than desired.

9 In an additional input receiving step 258, the tool 110 receives more text or menu
10 selections from a user to identify which of the context or contexts are desired for searching.
11 In an index reading step 260, the tool 110 reads an index of information contexts created in
12 conjunction with the database construction step 191. The relevant context or contexts in the
13 index are linked to site locations for information. The tool 110 returns these site locations
14 to a user in a pinpointed sites returning step 262.

15 Referring to Figure 11, the searching and filtering modules may alternatively be
16 embodied as shown in Figures 11 through 14, in contrast to the configurations shown in
17 Figures 5, 6, and 10. As above, the input module 116 may transmit text 117 reflecting a
18 search query to the filtering module 118, which may then filter information to isolate what a
19 user is seeking. In this embodiment, the filtering module 118 includes a context construction
20 module 300 for assembling micro-contexts 301 based on the text 117, a context comparison
21 module 302 for converting the micro-contexts 301 to macro-contexts 303, and an information
22 matching modules 304 for matching the macro-contexts 303 to specific information 306
23 responsive to the user's query. The presentation module 122 again provides the information
24 to a user.

25 The input module 116 may acquire text to describe information sought by a user in
26 a variety of different ways. For example, a simple free form text search may be used, wherein

1 the user types a query in plain language. Alternatively, a user may provide key words
2 separated by operators such as and, or, not, and others known in the art. The input module
3 116 may be configured to refine the text through questions to be answered by a user. The
4 filtering module 118 then receives the text from the input module 116. Until processed, the
5 text is only a series of words with no inherent meaning to a computer. The filtering module
6 118, in this embodiment, will convert the text into searchable portions to find matching
7 information of the type desired by a user.

8 Referring to Figure 12, the context construction module 300 is shown in greater
9 detail. The context construction module 300 assembles the words to form small, coherent
10 groups, or micro-contexts 301, for examples, they may contain about 1 to 5 words. This is
11 accomplished in part by using a block parser 316, which breaks down and interprets the text.
12 The text can be broken up by the block parser 300 in a number of different ways. Key words
13 312 and their modifiers, if designated by a user, can form or define natural contexts for
14 searching. Similarly, relative values 314 or priorities assigned to words in the text may be
15 used by the block parser 310 to create micro-contexts 301. Occurrence patterns 316 may be
16 used to form natural separations between groups of words.

17 These occurrence patterns 316 may be obtained from a user's history 318
18 corresponding to a given user's activities with the tool, including prior searches and results,
19 or from a general language database such as the attributes index 185. The user history 316
20 in any case may provide the tool 110 with information concerning what information a user has
21 requested in the past, and therefore what information the user is most likely looking for with
22 a new inquiry.

23 Referring to Figure 13, the context comparison module 302 is shown in greater
24 detail. The context comparison module 302 receives the micro-contexts 301 from the context
25 construction module 300 and compares them to a corpus 330 of information. The corpus 330
26 may simply be a database with samples of information 332 in natural language format, indexed

1 according to macro-contexts 303. These macro-contexts 303 may be more specific than the
2 micro-contexts 301.

3 The corpus 330 is sized to suit the amount and type of information on the network
4 10. The corpus 330, for example, may be composed of portions of text from 100,000 to
5 200,000 web sites, or more, with each portion matched to a macro-context 303. The entire
6 corpus 330 may be between 10 Megabytes and 10 Gigabytes in size, or larger. A text
7 comparison algorithm 336 may be provided to match text from the corpus 330 to the micro-
8 contexts 301, and then return the corresponding macro-context 303. The text comparison
9 algorithm 336 may combine several micro-contexts 301 to permit a more specific search,
10 thereby narrowing the number of matching macro-contexts 303.

11 Ideally, the context comparison module 302 will return a small number of macro-
12 contexts 303. However, this may not be possible for two reasons. First, if multiple,
13 important, micro-contexts 301 are identified by the context construction module 300, they
14 might not appear together within any portion of the corpus 330. In such a case, the context
15 comparison module 302 may return a series of macro-contexts 303 that match some fraction
16 of the important micro-contexts 301. Although these macro-contexts may not precisely
17 match a user's query, they may be ranked in order of likelihood that they will be relevant to
18 the user. The ranking may be obtained by using the user history 318 and other factors, such
19 as the number, probability, or nature of prior requests of the macro-context 303 by other
20 users, to determine the probability that a given macro-context 303 is relevant to the user.

21 Alternatively, the micro-contexts 301 may not even be found in the corpus 330. In
22 that case, a user may be referred to a user tracking module 338, which provides a user with
23 portals to access and search the network 10 directly. The user tracking module 338 permits
24 the tool 110 to track a user's progress through the network 10 to obtain further context
25 information for the current search, acquire more general information regarding contexts
26

1 important to the user, or find important information not currently present within the corpus
2 330.

3 A rapid mining module 339 may also be accessed while the user tracking module 338
4 is operating, to add nodes 52, or sites 52, to the corpus 330 and to process them through the
5 databasing module 114 "on the fly," or while the user is accessing the tool 110. These may
6 be sites 52 visited by the user or suggested by the user's query.

7 After searching the network 10 through the user tracking module 338, a user may
8 once again be referred to the input module 116 in order to provide additional text inputs, or
9 the context comparison module 302 may resume operation to process the micro-contexts 301
10 through new additions to the corpus 330.

11 Referring to Figure 14, the information matching module 304 receives macro-
12 contexts 303 from the context comparison module 302 and compares them to an indexed
13 database 350. The indexed database 350 contains specific information 306 of the type desired
14 by the user, indexed by macro-contexts 354 identical or similar to those provided by the
15 context comparison module 302. Thus, using a structure-matching algorithm 356, the
16 information matching module 304 can find the portion of specific information 306 that
17 correlates to the macro-contexts 303 provided by the context comparison module 302. The
18 specific information 306 located by the information matching module 304 may then be
19 returned to the presentation module 122 to be presented to a user.

20 The presentation module 122 is preferably flexible in its operation. For example, the
21 depth and breadth of specific information 306 returned may be varied according to a user's
22 preferences. Once the specific information 306 is located within the indexed database 350,
23 proximate information is easily gathered and returned. The order and arrangement of specific
24 information 306 displayed may also be determined manually by a user or automatically by
25 reference to the user history 318.
26

1 Referring to Figure 15, the indexed database 350 may be structured as a hierarchical
2 database 400. The hierarchical database 400 is configured like a tree, with general
3 information at the top and more specific information below. A parcel of information 402
4 desired by a user is a specific portion, and is therefore near the bottom of the hierarchical
5 database 400. According to traditional methods prior to the current invention, a user would
6 locate the parcel of information 402 by navigating through the broadest classification 404 and
7 through the branches 406, 408, and 410. A user might find this path difficult or even
8 impossible to find, particularly if the user knows little about the parcel of information 402, the
9 organizational scheme in which it resides, or the related elements in the hierarchy, and
10 therefore little about where it should be classified.

11 The current invention permits a user to navigate straight across hierarchies to the
12 parcel of information 402. The hierarchical database 400 remains transparent to the user,
13 who need not familiarize himself or herself with the structure of the hierarchical database 400.
14 Thus, the method disclosed herein provides horizontal navigation across a hierarchical
15 database, in which the tool 110 intelligently determines exactly what the user is looking for
16 and searches among the more specific, lower branches of the hierarchical database 400 to find
17 it.

18 One application of such a hierarchical, searchable database is to provide information
19 about products for sale over the Internet. In such a case, the presentation module 122
20 ultimately returns words to the user to denote the various products in the hierarchical
21 database 400 that match the user's request. The presentation module 122 may, for example,
22 be configured to sort products matching the user's request by brand, model, specifications,
23 price, merchant, availability, distance to the merchant from the user, shipping cost, or any
24 number of other relevant parameters.

25 In addition, the login module 166 may operate to navigate a site 52 for a user,
26 including forms presented by the site 52 to collect information from the user. Thus, not only

1 is a user freed from the need to navigate the hierarchical database, the user may also be
2 permitted to access the site 52 and conduct business on it without having to navigate the
3 structure of the site 52.

4 The tool 110 as configured above is also well adapted for use without such a
5 hierarchical structure. The context matching capabilities of the tool 110 make the tool 110
6 effective for improving the relevance and completeness of results to a query, regardless of
7 what formats are used by the tool 110 to maintain and organize a local database. This is a
8 vast improvement over current search engines, which typically search only for the exact text
9 provided by the user, and thus deliver results that include irrelevant items and fail to include
10 important information.

11 Referring to Figure 16, an alternative and more detailed embodiment of the user
12 interface 134 is shown. A user may first be directed to a query input page 500, in which the
13 user may provide text descriptive of the desired information in a text input module 502. The
14 text input module 502 may receive text in natural language form, i.e. as a sentence or
15 question, or may be configured to accept keywords separated by operators, such as "and,"
16 "or," and "not." A parameter selection module 504 permits a user to adjust various
17 parameters for the search from the outset, such as the amount and type of results to display.
18 A user is not required to select categories of the information desired through menu structures.

19 Once a user has provided the text, the filtering module 118 operates substantially in
20 the background, such that a user is not aware of its operation. However, as the filtering
21 module 118 identifies relevant contexts, multiple potentially relevant macro-contexts 303 may
22 be identified. Thus, the tool 110 may transmit a context refinement page 510 to a user, in
23 which a context inquiry module 512 permits a user to select which macro-context 303 is most
24 relevant. The tool would then transmit an information presentation page 520 containing
25 information about the chosen macro-context 303.
26

1 Alternatively, the tool 110 may take a user directly from the query input page 500
2 to the information presentation page 520. The information presentation page 520, as
3 embodied in Figure 16, contains an information chart 522 with product information 524 of
4 the type desired by a user, as well as product locations 526 showing where a product may be
5 purchased on the Internet. The product information 524 may include items relevant to a user,
6 such as the manufacturer, model, SKU number, price, availability, shipping cost, and relevant
7 characteristics of each product. The product locations 526 may take the form of links to take
8 a user directly to a site where a given product may be purchased.

9 Additionally, the information presentation page 520 may contain presentation
10 parameters 526 that can be adjusted by a user. For example, a user may alter the context to
11 select different products with a context selector 530, adjust the depth to include more
12 products in the search with a depth selector 532, increase the breadth of the search to include
13 a wider variety of products with a breadth selector 534, alter the order in which products are
14 presented in the information chart 522 with an order selector 536, and more specifically define
15 which products are shown in the information chart 522 with a filtering selector 538.

16 It is possible that the tool 110 may not find any products matching a user's query.
17 In that case, the tool 110 may transmit a user search page 550, which permits a user to search
18 the Internet directly from within the user interface 134. An additional text input module 552
19 may be provided to permit a user to input different text for a general search of the Internet.
20 As the tool 110 searches the Internet, the tracking module 338 operates in the background
21 to follow and gather information from a user's navigation. Thus, the tracking module 338
22 does not appear as part of the user interface 134. After a user searches the Internet, the tool
23 110 may take a user to the information presentation page 520 to present information gathered
24 by the tracking module 338.

25 When a user finds a product and merchant of interest, he or she selects the
26 appropriate product location 526, and the tool 110 may send a user to a merchant's site 560,

1 where a user can obtain more information about the product and order it if desired. The
2 merchant's site 560 may also be displayed within the user interface 134 of the tool 110, so
3 that a user has not truly left the tool 110. The tool 110 may be configured to transmit the
4 appropriate user information to the merchant's site 560 to "log in" a user and obtain access
5 to information without requiring a user to fill out additional forms from the merchant's site
6 560.

7 Alternatively, the tool 110 may, itself, store additional information about the product
8 and transmit it in the form of a product presentation page 570. The product presentation
9 page 570 contains product details 572 and a purchase option 574, whereby a user may learn
10 more about the product and purchase it as though at the merchant's site 560.

11 When a user chooses to purchase a product, either from the merchant's site 560
12 viewed through the user interface 134 of the tool, or from the product presentation page 570,
13 using the purchase option 574, the tool 110 may respond by transmitting a purchase page
14 580. The purchase page 580 provides or obtains from a user additional information necessary
15 to complete the transaction, such as shipping addresses and credit card numbers.
16 Alternatively, the tool 110 may be configured to transmit this information automatically to
17 the merchant's site 560. In any case, products are ordered through the merchants' site 560,
18 whether through the purchase page 580 provided by the tool 110, or through the merchant's
19 site 560 itself. A user may select, from the information presentation page 520, the merchant's
20 site 560, the product presentation page 570, or the purchase page 580, to return to the query
21 input page 500 for a new search.

22 Referring to Figure 17, an exemplary information chart 522 is provided. If a user
23 were to submit a search for sleeping bags, results of the search may be similar to those shown
24 in Figure 17. The information chart 522 contains product information 606 arrayed with each
25 product occupying a single row. Various attributes of each product are grouped by column,
26 including the merchant 610 offering the product for sale, manufacturer 620, model 630,

1 physical attributes 640 of the product, and price 670 of the product. The physical attributes
2 640 of the product are items of interest to a consumer, such as the maximum height 650 of
3 a person able to use the sleeping bag, the loft 652, or thickness, of the bag, the weight 654
4 of the bag, and the lowest external temperature 656 for comfortable sleeping in the bag.

5 Since different users will be interested in different characteristics, the tool 110
6 provides the presentation parameters 528 to allow a user to alter what is shown in the
7 information chart 522. For example, if a user were to decide that he or she would prefer an
8 outdoor blanket, the context selector 530 may be used to effectively change the search to
9 display similar results for outdoor blankets. A user may also use the depth selector 532 to
10 show only broader product categories, such as manufacturers, product lines, or physical
11 attributes 640. A user may, for example, choose to display only lowest external temperatures
12 656, to see the temperatures available without having to sort through several bags with the
13 same lowest external temperature 656.

14 With the breadth selector 534, a user may choose to broaden or narrow the search,
15 for example, to show both sleeping bags and outdoor blankets, or to show only sleeping bags
16 from a certain manufacturer. The filtering selector 538 may offer a higher degree of control
17 over what is displayed by allowing a user to select to include or exclude specific products or
18 characteristics. With the order selector 536, a user may modify the order of results shown,
19 for example, to show cheaper sleeping bags first, or to display lighter bags before heavier
20 ones. The order selector 536 may permit a user to order by one attribute, then by another
21 when the first attributes are equal. Thus, a user has full control over what the information
22 chart 522 shows and how the information is grouped in the chart 522. A user can make a
23 better and more rapid decision by discarding irrelevant information and properly arranging the
24 remaining products.

25 From the above discussion, it will be appreciated that the present invention provides
26 a data extraction tool for extracting information from an information source. Extracted

1 information is cataloged and indexed for future searching by a user. Although not limited to
2 commerce, the method disclosed herein may be adapted to search for commerce-ready web
3 sites on the Internet.

4 The present invention may be embodied in other specific forms without departing
5 from its structures, methods, or other essential characteristics as broadly described herein and
6 claimed hereinafter. The described embodiments are to be considered in all respects only as
7 illustrative, and not restrictive. The scope of the invention is, therefore, indicated by the
8 appended claims, rather than by the foregoing description. All changes that come within the
9 meaning and range of equivalency of the claims are to be embraced within their scope.

10 What is claimed and desired to be secured by United States Letters Patent is:
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26